Solution 1: High-dimensional Gaussian Distributions

We will use that $\|X\|_2^2$ is in fact a sum of squared Gaussian random variables:

$$\|X\|_2^2 = \sum_{i=1}^{p} X_i^2. \tag{S}$$

(a) Using (S) we derive directly

$$\mathbb{E}(\|X\|_2^2) = \mathbb{E}\left(\sum_{i=1}^{p} X_i^2\right) = \sum_{i=1}^{p} \mathbb{E}X_i^2 = \sum_{i=1}^{p} 1 = p, \tag{E}$$

since $X_i \sim \mathcal{N}(0,1)$ for each $i = 1, \dots, p$. Again using (S) we obtain

$$\begin{aligned}
\mathsf{Var}(\|X\|_2^2) &= \mathsf{Var}\left(\sum_{i=1}^{p} X_i^2\right) && \text{(Using (S))} \\
&= \sum_{i=1}^{p} \mathsf{Var}\left(X_i^2\right) && (X_1, \dots, X_p \text{ are i.i.d.}) \\
&= \sum_{i=1}^{p} (\mathbb{E}(X_i^4) - \mathbb{E}(X_i^2)^2) && (\mathsf{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 \text{ for any rv } Y) \\
&= \sum_{i=1}^{p} (3-1) && (\text{Using the hint: } \mathbb{E}_{Y\sim\mathcal{N}(0,1)}(Y^4) = 3) \\
&= 2p.
\end{aligned}$$

(b) We write

$$\|X\|_2 - \sqrt{p} = \underbrace{\frac{\|X\|_2^2 - p}{2\sqrt{p}}}_{=(1)} - \underbrace{\frac{(\|X\|_2^2 - p)^2}{2\sqrt{p}(\|X\|_2 + \sqrt{p})^2}}_{=(2)}.$$

It holds that

$$\begin{aligned}
\mathbb{E}[(1)] &= \mathbb{E}\left[\frac{\|X\|_2^2 - p}{2\sqrt{p}}\right] \\
&= \frac{1}{2\sqrt{p}}\left(\mathbb{E}\left[\|X\|_2^2\right] - p\right) && \text{(Linearity of expected value)} \\
&= \frac{1}{2\sqrt{p}}(p - p) && \text{(Using (E))} \\
&= 0.
\end{aligned}$$

On the one hand, it holds that $0 \le (2)$, as all terms are non-negative and consequently $0 \le \mathbb{E}[(2)]$. On the other hand, since $\|X\|_2 \ge 0$ we have

$$\begin{aligned}
& (2) \le \frac{(\|X\|_2^2 - p)^2}{2p^{3/2}} \\
\Rightarrow \quad & \mathbb{E}[(2)] \le \mathbb{E}\left[\frac{(\|X\|_2^2 - p)^2}{2p^{3/2}}\right] = \frac{\mathsf{Var}(\|X\|_2^2)}{2p^{3/2}} = \frac{1}{\sqrt{p}}.
\end{aligned}$$

Putting everything together:

$$|\mathbb{E}(\|X\|_2 - \sqrt{p})| = |\underbrace{\mathbb{E}[(1)]}_{=0} - \mathbb{E}[(2)]| = \mathbb{E}[(2)] \le \frac{1}{\sqrt{p}}. \tag{b}$$

(c) The variance can be bounded as follows:

$$\begin{aligned}
\mathsf{Var}\left(\|X\|_2\right) &= \mathsf{Var}\left(\|X\|_2 - \sqrt{p}\right) && \text{(Variance does not change by constant shifts)}\\
&\le \mathbb{E}\left[(\|X\|_2 - \sqrt{p})^2\right] && \text{(For any random variable } Y \text{ it holds that } \mathsf{Var}(Y) \le \mathbb{E}(Y^2))\\
&= \mathbb{E}\left[\|X\|_2^2 - 2\sqrt{p}\|X\|_2 + p\right]\\
&= \mathbb{E}\left[\|X\|_2^2\right] - 2\sqrt{p}\mathbb{E}\left[\|X\|_2\right] + p && \text{(Linearity of expected value)}\\
&= 2p - 2\sqrt{p}\mathbb{E}\left[\|X\|_2\right] && \text{(Using (E))}\\
&= 2p - 2\sqrt{p}\mathbb{E}[\|X\|_2 - \sqrt{p} + \sqrt{p}] && \text{(Someone told us that it is a good idea)}\\
&= -2\sqrt{p}\mathbb{E}[\|X\|_2 - \sqrt{p}] && \text{(Linearity of expected value)}\\
&\le 2\sqrt{p}\frac{1}{\sqrt{p}} = 2. && \text{(Using (b))}
\end{aligned}$$

(d) Since $Z = \frac{X-X'}{\sqrt{2}} \sim \mathcal{N}(0, \boldsymbol{I})$, we obtain from (b) and (c) that

$$\left|\mathbb{E}(\|Z\|_2 - \sqrt{p}\ )\right| \le \sqrt{\frac{1}{p}}, \tag{d1}$$

$$\mathsf{Var}\left(\|Z\|_2\right) \le 2. \tag{d2}$$

But

$$\|Z\|_2 = \sqrt{\sum_{i=1}^{p}\left(\frac{X_i - X_i'}{\sqrt{2}}\right)^2} = \sqrt{\frac{1}{2}\sum_{i=1}^{p}(X_i - X_i')^2} = \sqrt{\frac{1}{2}}\sqrt{\sum_{i=1}^{p}(X_i - X_i')^2} = \sqrt{\frac{1}{2}}\|X - X'\|_2. \tag{d3}$$

Thus, (d1) implies

$$\begin{aligned}
&\left|\mathbb{E}(\|Z\|_2 - \sqrt{p}\ )\right| \le \sqrt{\frac{1}{p}}\\
\Leftrightarrow\quad &\sqrt{2}\left|\mathbb{E}(\|Z\|_2 - \sqrt{p}\ )\right| \le \sqrt{\frac{2}{p}}\\
\Leftrightarrow\quad &\Big|\mathbb{E}(\underbrace{\sqrt{2}\|Z\|_2}_{=\|X-X'\|_2} - \sqrt{2p}\ )\Big| \le \sqrt{\frac{2}{p}}.
\end{aligned}$$

Moreover, (d2) implies

$$\begin{aligned}
&\mathsf{Var}\left(\|Z\|_2\right) \le 2\\
\Leftrightarrow\quad &2\mathsf{Var}\left(\|Z\|_2\right) \le 2 \cdot 2\\
\Leftrightarrow\quad &\mathsf{Var}\left(\sqrt{2}\|Z\|_2\right) \le 4 && \text{(}\mathsf{Var}(aY) = a^2\mathsf{Var}(Y) \text{ for any rv } Y \text{ and constant } a\text{)}\\
\Leftrightarrow\quad &\mathsf{Var}\left(\|X - X'\|_2\right) \le 4. && \text{(Using (d3))}
\end{aligned}$$

(e) As for any $x, x' \in \mathbb{R}^p$ it holds that

$$\langle x, x'\rangle = \frac{1}{2}(\|x\|_2^2 + \|x'\|_2^2 - \|x - x'\|_2^2).$$

we can infer that

$$\begin{aligned}
\mathbb{E}\langle X, X'\rangle &= \frac{1}{2}\left(\mathbb{E}\|X\|_2^2 + \mathbb{E}\|X'\|_2^2 - \mathbb{E}\|X - X'\|_2^2\right)\\
&= \frac{1}{2}\left(p + p - \mathbb{E}\|X - X'\|_2^2\right) && \text{(Using (E))}\\
&= \frac{1}{2}\left(p + p - 2\mathbb{E}\underbrace{\frac{1}{2}\|X - X'\|_2^2}_{=\|Z\|_2^2}\right) && \text{(Using (d3))}\\
&= \frac{1}{2}(p + p - 2p) = 0. && \text{(Using again (E))}
\end{aligned}$$

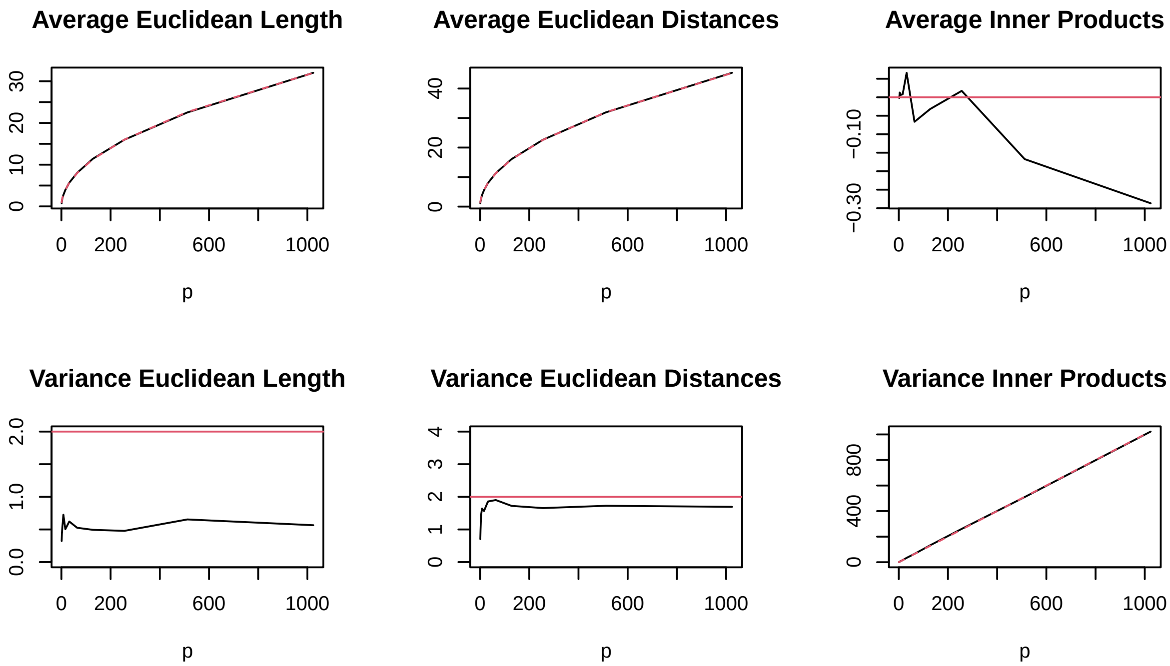
For the variance we obtain

$$
\begin{aligned}
\mathsf{Var}(\langle X, X'\rangle) &= \mathsf{Var}\left(\sum_{i=1}^{p} X_i\, X_i'\right) \\
&= \sum_{i=1}^{p} \mathsf{Var}\left(X_i\, X_i'\right) && \text{(Independence)} \\
&= p\,\mathsf{Var}\left(X_1\, X_1'\right) && \text{(Identical distributions)} \\
&= p\, \left(\mathbb{E}[X_1^2\,(X_1')^2] - \mathbb{E}[X_1\,(X_1')]^2\right) \\
&= p\, \left(\mathbb{E}[X_1^2]\mathbb{E}[(X_1')^2] - \mathbb{E}[X_1]^2\,\mathbb{E}[(X_1')]^2\right) && \text{(Independence)} \\
&= p. && (\mathbb{E}[X_1^2] = \mathbb{E}[(X_1')^2] = 1 \text{ and } \mathbb{E}[X_1] = \mathbb{E}[X_1'] = 0)
\end{aligned}
$$

(f)

```
# load library to sample from multivariate normal distribution
library(mvtnorm)

# compute average euclidean length of a matrix x (rows = samples)
average_euclidean_length <- function(x){
  mean(apply(x,1,norm,type="2"))
}

# compute variance of euclidean lengths of a matrix x (rows = samples)
variance_euclidean_length <- function(x){
  var(apply(x,1,norm,type="2"))
}

# compute average euclidean distances between matrices x and x2 (rows = samples)
average_euclidean_distances <- function(x,x2){
  z = c()
  for (i in 1:nrow(x)){
    z = rbind(z,x[i,]-x2)
  }
  mean(apply(z,1,norm,type="2"))
}

# compute variance of euclidean distances between matrices x and x2
variance_euclidean_distances <- function(x,x2){
  z = c()
  for (i in 1:nrow(x)){
    z = rbind(z,x[i,]-x2)
  }
  var(apply(z,1,norm,type="2"))
}

# compute average inner products between matrices x and x2 (rows = samples)
average_inner_product <- function(x,x2){
  z = c()
  for (i in 1:nrow(x)){
    z = rbind(z,x2%*%x[i,])
  }
  mean(z)
}

# compute varaince of inner products between matrices x and x2 (rows = samples)
variance_inner_product <- function(x,x2){
  z = c()
  for (i in 1:nrow(x)){
    z = rbind(z,x2%*%x[i,])
```

```
  }
  var(z)
}


set.seed(5)

p_range         <- 2^seq(0,10)

avg_eucl_length <- c()
var_eucl_length <- c()
avg_eucl_dist   <- c()
var_eucl_dist   <- c()
avg_inner_prod  <- c()
var_inner_prod  <- c()

for (p in p_range){

  x               <- rmvnorm(n=100, mean=rep(0,p), sigma=diag(p))
  x2              <- rmvnorm(n=100, mean=rep(0,p), sigma=diag(p))

  avg_eucl_length <- c(avg_eucl_length,average_euclidean_length(x))
  var_eucl_length <- c(var_eucl_length,variance_euclidean_length(x))
  avg_eucl_dist   <- c(avg_eucl_dist,average_euclidean_distances(x,x2))
  var_eucl_dist   <- c(var_eucl_dist,variance_euclidean_distances(x,x2))
  avg_inner_prod  <- c(avg_inner_prod,average_inner_product(x,x2))
  var_inner_prod  <- c(var_inner_prod,variance_inner_product(x,x2))


}

# compare the results visually
par(mfrow=c(2,3))

plot(p_range,avg_eucl_length,type="l",main="Average Euclidean Length",xlab="p",ylab="")
lines(p_range,sqrt(p_range),col=2,lty=2)

plot(p_range,avg_eucl_dist,type="l",main="Average Euclidean Distances",xlab="p",ylab="")
lines(p_range,sqrt(2*p_range),col=2,lty=2)

plot(p_range,avg_inner_prod,type="l",main="Average Inner Products",xlab="p",ylab="")
abline(h=0,col=2)

plot(p_range,var_eucl_length,type="l",main="Variance Euclidean Length",xlab="p",ylab="",ylim=c(0,2))
abline(h=2,col=2)

plot(p_range,var_eucl_dist,type="l",main="Variance Euclidean Distances",xlab="p",ylab="",ylim=c(0,4)
abline(h=2,col=2)

plot(p_range,var_inner_prod,type="l",main="Variance Inner Products",xlab="p",ylab="")
lines(p_range,p_range,col=2,lty=2)
```


Average Euclidean Length
Average Euclidean Distances
Average Inner Products
Variance Euclidean Length
Variance Euclidean Distances
Variance Inner Products
p